CRYSTALLIZATION AND STRUCTURE DETERMINATION OF *STAPHYLOCOCCUS AUREUS* THYMIDYLATE KINASE

This application claims the benefit of U.S. Provisional Application Serial No. 60/147,117, filed 4 August 1999, which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

This invention relates to the crystallization and structure determination of thymidylate kinase (TMK) from *Staphylococcus aureus*.

BACKGROUND

Thymidylate kinase (TMK) catalyzes the synthesis of (deoxy)thymidine diphosphate (dTDP) from (deoxy)thymidine monophosphate (dTMP) and ATP along the pathway leading to the synthesis of (deoxy)thymidine triphosphate (dTTP) necessary for DNA synthesis (Figure 1). Since the phosphorylation of dTDP to dTTP is conducted by a nonspecific diphosphate kinase, TMK is a key player in the regulation of DNA synthesis and is a potential antibacterial target. Interest in thymidylate kinase biochemistry increased when it was recently discovered that this enzyme serves as one of the activators for the AIDS drug, 3'-azido-3'-deoxythymidine (AZT) (L.W. Frick et al., Biochem. Biophys. Res. Comm. 154:124-9 (1988); A. Fridland et al., Mol. Pharmacol. 37:665-70 (1990)). Activation of AZT to azidothymidine triphosphate (AZT-TP) proceeds along cellular phosphorylation pathways to produce the species which is incorporated into growing DNA chains by HIV reverse transcriptase. Similar to its role in serving as a control point for the production of dTTP, thymidylate kinase catalyzes the rate limiting phosphorylation of AZT-monophosphate to AZT-diphosphate (AZT-DP). AZT-DP phosphorylation to AZT-TP is then catalyzed by a nonspecific diphosphate kinase.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides a method for crystallizing an *S. aureus* thymidylate kinase molecule or molecular complex that includes preparing purified *S. aureus* thymidylate kinase at a concentration of about 1 mg/ml to about 50 mg/ml and crystallizing *S. aureus* thymidylate kinase from a solution including about 5 wt. % to about 50 wt. % PEG (preferably having a number average molecular weight between about 200 and about 20,000), about 0.05 M to about 0.5 M $MgCl_2$, and about 0 wt. % to about 20 wt. % DMSO, wherein the solution is buffered to a pH of about 6 to about 7. In another aspect, the present invention provides a method for crystallizing an *S. aureus* thymidylate kinase molecule or molecular complex that includes preparing purified *S. aureus* thymidylate kinase at a concentration of about 1 mg/ml to about 50 mg/ml and crystallizing *S. aureus* thymidylate kinase from a solution including about 2 mM to about 20 mM β,γ-difluoromethylene-bisphosphonate adenosine monophosphate and about 0 wt. % to about 20 wt. % DMSO, wherein the solution is buffered to a pH of about 6 to about 7

In another aspect, the present invention provides crystalline forms of an *S. aureus* thymidylate kinase molecule. In one embodiment, a crystal of *S. aureus* thymidylate kinase is provided having the trigonal space group symmetry $P2_1$.

In another aspect, the present invention provides a scalable three dimensional configuration of points derived from structure coordinates of at least a portion of an *S. aureus* thymidylate kinase molecule or molecular complex. In one embodiment, the scalable three dimensional set of points is derived from structure coordinates of at least the backbone atoms of the amino acids representing a TMP and/or TMP/ATP substrate binding pocket of an *S. aureus* thymidylate kinase molecule or molecular complex. In another embodiment, the scalable three dimensional configuration of points is derived from structure coordinates of at least

a portion of a molecule or a molecular complex that is structurally homologous to an *S. aureus* thymidylate kinase molecule or molecular complex. On a molecular scale, the configuration of points derived from a homologous molecule or molecular complex have a root mean square deviation of less than about 2.1 Å from the structure coordinates of the molecule or complex

In another aspect, the present invention provides a molecule or molecular complex that includes at least a portion of an *S. aureus* thymidylate kinase TMP and/or TMP/ATP substrate binding pocket. In one embodiment, the *S. aureus* thymidylate kinase TMP substrate binding pocket includes the amino acids listed in Table 1, preferably the amino acids listed in Table 2, and more preferably the amino acids listed in Table 3, the substrate binding pocket being defined by a set of points having a root mean square deviation of less than about 2.1 Å, preferably less than about 1.5 Å, more preferably less than about 1.0 Å, and most preferably less than about 0.5 Å from points representing the backbone atoms of the amino acids. In another embodiment, the *S. aureus* thymidylate kinase TMP/ATP substrate binding pocket includes the amino acids listed in Table 4, preferably the amino acids listed in Table 5, and more preferably the amino acids listed in Table 6, the substrate binding pocket being defined by a set of points having a root mean square deviation of less than about 2.1 Å, preferably less than about 1.5 Å, more preferably less than about 1.0 Å, and most preferably less than about 0.5 Å from points representing the backbone atoms of the amino acids.

Table 1

Residues within about 4Å of the TMP binding pocket of *S. aureus* TMK

GLU	12	LEU	53	ARG	93
ARG	37	LEU	66	SER	97
ILE	48	PHE	67	SER	98
ARG	49	SER	70	TYR	101
VAL	52	ARG	71		

Table 2
Residues within about 7Å of the TMP binding pocket of *S. aureus* TMK

GLY	10	VAL	52	TYR	94
GLU	12	LEU	53	ILE	95
ARG	37	GLU	63	ASP	96
GLU	38	MET	65	SER	97
PRO	39	LEU	66	SER	98
GLY	45	PHE	67	LEU	99
GLU	46	ALA	68	ALA	100
GLU	47	ALA	69	TYR	101
ILE	48	SER	70	GLN	102
ARG	49	ARG	71	ASN	117
LYS	50	ASP	92	PHE	160
ILE	51	ARG	93	TYR	168

Table 3
Residues within about 10Å of the TMP binding pocket of *S. aureus* TMK

PHE	8	ILE	51	TYR	94
GLU	9	VAL	52	ILE	95
GLY	10	LEU	53	ASP	96
PRO	11	GLU	54	SER	97
GLU	12	GLY	55	SER	98
GLY	13	MET	58	LEU	99
SER	14	ILE	60	ALA	100
LYS	16	THR	62	TYR	101
THR	17	GLU	63	GLN	102
ARG	37	ALA	64	GLY	103
GLU	38	MET	65	TYR	104
PRO	39	LEU	66	ALA	105
GLY	40	PHE	67	ARG	106
GLY	41	ALA	68	VAL	113
VAL	42	ALA	69	LEU	116
PRO	43	SER	70	ASN	117
THR	44	ARG	71	ILE	143
GLY	45	ARG	72	PHE	160
GLU	46	GLU	73	HIS	161
GLU	47	HIS	74	VAL	164
ILE	48	CYS	91	TYR	168
ARG	49	ASP	92		
LYS	50	ARG	93		

Table 4

Residues within about 4Å of the TMP/ATP binding pocket of *S. aureus* TMK

GLU	12	GLU	38	SER	98
GLY	15	PHE	67	TYR	101
LYS	16	ARG	71	GLN	102
THR	17	ASP	92	ARG	142
THR	18	ARG	93	LEU	188
ARG	37	SER	97		

Table 5

Residues within about 7Å of the TMP/ATP binding pocket of *S. aureus* TMK

		ARG	49	TYR	101
GLY	10	GLU	63	GLN	102
PRO	11	ALA	64	ARG	106
GLU	12	PHE	67	ASN	117
GLY	13	ALA	68	LEU	132
SER	14	ARG	71	GLU	141
GLY	15	ASP	92	ARG	142
LYS	16	ARG	93	ILE	143
THR	17	TYR	94	PHE	160
THR	18	ILE	95	ALA	184
VAL	19	ASP	96	GLN	186
ILE	20	SER	97	PRO	187
ASN	21	SER	98	LEU	188
ARG	37	LEU	99	GLU	189
GLU	38	ALA	100	VAL	191

Table 6

Residues within about 10Å of the TMP/ATP binding pocket of *S. aureus* TMK

PHE	8	ALA	64	VAL	113
GLU	9	MET	65	LEU	116
GLY	10	LEU	66	ASN	117
PRO	11	PHE	67	ALA	120
GLU	12	ALA	68	LEU	132
GLY	13	ALA	69	VAL	134
SER	14	SER	70	VAL	138
GLY	15	ARG	71	GLY	139
LYS	16	ARG	72	ARG	140
THR	17	HIS	74	GLU	141
THR	18	CYS	91	ARG	142
VAL	19	ASP	92	ILE	143
ILE	20	ARG	93	ASP	157
ASN	21	TYR	94	PHE	160
GLU	22	ILE	95	HIS	161
MET	35	ASP	96	VAL	164
THR	36	SER	97	TYR	168
ARG	37	SER	98	ASN	183
GLU	38	LEU	99	ALA	184
PRO	39	ALA	100	ASP	185
GLY	40	TYR	101	GLN	186
GLU	46	GLN	102	PRO	187
ARG	49	GLY	103	LEU	188
VAL	52	TYR	104	GLU	189
LEU	53	ALA	105	ASN	190
ILE	60	ARG	106	VAL	191
GLU	63	ILE	108	VAL	192

In another aspect, the present invention provides molecules or molecular complexes that are structurally homologous to an *S. aureus* thymidylate kinase molecule or molecular complex.

In another aspect, the present invention provides a machine readable storage medium including the structure coordinates of all or a portion of an *S. aureus* thymidylate kinase molecule, molecular complex, a structurally homologous molecule or complex, including structurally equivalent structures, as defined herein, particularly a substrate binding pocket thereof, or a similarly shaped homologous

substrate binding pocket. A storage medium encoded with these data is capable of displaying on a computer screen, or similar viewing device, a three-dimensional graphical representation of a molecule or molecular complex which comprises a substrate binding pocket or a similarly shaped homologous substrate binding pocket.

In another aspect, the present invention provides a method for identifying inhibitors, ligands, and the like for an *S. aureus* thymidylate kinase molecule by providing the coordinates of a molecule of *S. aureus* thymidylate kinase to a computerized modeling system; identifying chemical entities that are likely to bind to or interfere with the molecule (e.g., screening a small molecule library); and, optionally, procuring or synthesizing and assaying the compounds or analogues derived therefrom for bioactivity. In another aspect, the present invention provides methods for designing inhibitors, ligands, and the like by providing the coordinates of a molecule of *S. aureus* thymidylate kinase to a computerized modeling system; designing a chemical entity that is likely to bind to or interfere with the molecule; and optionally synthesizing the chemical entity and assaying the chemical entity for bioactivity. In another aspect, the present invention provides inhibitors and ligands designed or identified by the above methods. In one embodiment, a composition is provided that includes an inhibitor or ligand designed or identified by the above method. In another embodiment, the composition is a pharmaceutical composition.

In another aspect, the present invention provides a method involving molecular replacement to obtain structural information about a molecule or molecular complex of unknown structure. The method includes crystallizing the molecule or molecular complex, generating an x-ray diffraction pattern from the crystallized molecule or molecular complex, and applying at least a portion of the structure coordinates set forth in Fig. 2 to the x-ray diffraction pattern to generate a three-dimensional electron density map of at least a portion of the molecule or molecular complex.

In another aspect, the present invention provides a method for homology modeling an *S. aureus* thymidylate kinase homolog.

DEFINITIONS

Two crystallographic data sets (with structure factors F) are considered isomorphous if, after scaling,

$$\frac{\Delta F}{F} = \frac{\sum |F_1 - F_2|}{\sum F_1}$$

is less than about 35% for the reflections between 8 Å and 4 Å.

ABBREVIATIONS

The following abbreviations are used throughout this disclosure:

Staphylococcus aureus (*S. aureus*).

Thymidylate kinase (T. kinase or TMK).

Thymidine 5'-monophosphate (TMP).

Thymidine 5'-diphosphate (TDP).

Thymidine 5'-triphosphate (TTP).

Phospho(enol)pyruvate (PEP)

Reduced nicotinamide adenine dinucleotide (NADH)

Oxidized nicotinamide adenine dinucleotide (NAD^+)

Pyruvate kinase (PK)

Lactate dehydrogenase (LDH)

Nucleoside-5'-diphosphate kinase (NDP-Kinase)

(Deoxy)thymidine monophosphate (dTMP).

(Deoxy)thymidine diphosphate (dTDP).

(Deoxy)thymidine triphosphate (dTTP).

Adenosine 5'-diphosphate (ADP).

Adenosine 5'-triphosphate (ATP).

Isopropylthiò-β-D-galactoside (IPTG).

Dithiothreitol (DTT).

Dimethyl sulfoxide (DMSO).

Polyethylene glycol (PEG).

Multiple anomalous dispersion (MAD).

The following amino acid abbreviations are used throughout this disclosure:

A = Ala = Alanine	T = Thr = Threonine
V = Val = Valine	C = Cys = Cysteine
L = Leu = Leucine	Y = Tyr = Tyrosine
I = Ile = Isoleucine	N = Asn = Asparagine
P = Pro = Proline	Q = Gln = Glutamine
F = Phe = Phenylalanine	D = Asp = Aspartic Acid
W = Trp = Tryptophan	E = Glu = Glutamic Acid
M = Met = Methionine	K = Lys = Lysine
G = Gly = Glycine	R = Arg = Arginine
S = Ser = Serine	H = His = Histidine

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the biosynthetic pathway for the synthesis of thymidylate. The reaction catalyzed by thymidylate kinase is boxed.

Figure 2 lists the atomic structure coordinates for recombinant *S. aureus* thymidylate kinase (with a His_6 tag) as derived by x-ray diffraction from a crystal of that complex. The following abbreviations are used in Figure 2:

"Atom" refers to the element whose coordinates are measured. The second column defines the number of the atom in the structure. The letters in the

third column define the element. The fourth and fifth columns define the amino acid and the number of the amino acid in the structure, respectively.

"X, Y, Z" crystallographically define the atomic position of the element measured.

"Occ" is an occupancy factor that refers to the fraction of the molecules in which each atom occupies the position specified by the coordinates. A value of "1" indicates that each atom has the same conformation, i.e., the same position, in all molecules of the crystal.

"B" is a thermal factor that measures movement of the atom around its atomic center.

Figure 3 depicts *S. aureus* thymidylate kinase using (a) a ribbon diagram showing the backbone structure of the enzyme and (b) a schematic diagram showing the secondary structure for a TMK monomer Disordered loops are indicated by arrows.

Figure 4 depicts a structural comparison of *E. coli* TMK + AP_5T and *S. aureus* TMK. The overall fold of the two proteins is well-conserved, but note that the lid in the *E. coli* TMK is not present in the *S. aureus* TMK due to the absence of a ligand.

Figure 5 depicts (a) a stereo view of a superposition of *S. aureus* thymidylate kinase and *E. coli* thymidylate kinase and (b) the amino acid sequence alignment of *S. aureus* thymidylate kinase (SEQ ID NO:1) (capital letters, upper sequence) and *E. coli* thymidylate kinase (SEQ ID NO:2) (lower sequence). Dots in the sequences indicate gaps inserted in order to optimize the alignment. Identical residues are indicated by | and similar residues are indicated by . and : symbols.

Figure 6 depicts (a) a stereo view of a superposition of *S. aureus* thymidylate kinase and *S. cerevisiae* thymidylate kinase and (b) the sequence alignment of *S. aureus* thymidylate kinase (SEQ ID NO:1) (capital letters, upper sequence) and *S. cerevisiae* thymidylate kinase (SEQ ID NO:3) (lower sequence). Dots in the sequences indicate gaps inserted in order to optimize the alignment.

Identical residues are indicated by | and similar residues are indicated by . and : symbols.

Figure 7 depicts a) a substrate-based inhibitor (AP_5T) for thymidylate kinase with a K_d of 20 nM for *E. coli* TMK (A. Lavie et al., Biochemistry 37:3677-86 (1998); A. Lavie et al., Proc. Natl. Acad. Sci. USA, 95:14045-50 (1998)). b) protein ligand interactions for *E. coli* TMK (shaded boxes, from A. Lavie et al., Proc. Natl. Acad. Sci. USA, 95:14045-50 (1998)) with the corresponding residues from *S. aureus* TMK underlined (conservative mutations are marked with an asterisk). Active site residues from the *S. cerevisiae* are boxed (where no corresponding residue from *E. coli* TMK is present, an arrow indicates the point of contact with the substrate).

Figure 8 depicts the anomalous difference Patterson maps at (a) 2.7 Å and (b) at 2.3 Å resolution.

Figure 9 depicts electron density maps of residues 76 to 82 from molecule 1 of *S. aureus* thymidylate kinase (SEQ ID NO:1) at (a) 2.7 Å and (b) at 2.3 Å resolution.

Figure 10 lists the structure factors and multiple anomalous dispersion phases for the crystal structure of *S. aureus* thymidylate kinase (SEQ ID NO:1). "INDE" refers to the indices h, k, and l (columns 2, 3, and 4 respectively) of the lattice planes. "FOBS" refers to the structure factor (F) of the observed reflections. "SIGMA" is the standard deviation for the observations. "PHAS" refers to the phase used for the observations. "FOM" refers to the figure of merit.

Figure 11 depicts a surface representation of a) *E. coli* TMK with the inhibitor AP_5T and b) *S. aureus* TMK with a hypothetical positioning of AP_5T based on a structural alignment of C_α atoms from the *E. coli* TMK + AP_5T structure.

DETAILED DESCRIPTION OF THE INVENTION

Crystalline Form(s) and Method of Making

The three-dimensional structure of *S. aureus* thymidylate kinase was solved using high resolution x-ray crystallography to 2.3 Å resolution (see Figure 2 and Example 1). Accordingly, the invention includes a TMK crystal and/or a crystal with TMK co-crystallized with a ligand, such as an inhibitor. Preferably, the crystal has trigonal space group symmetry $P2_1$. More preferably, the crystal comprises rectangular shaped unit cells, each unit cell having dimensions of a, b, and c; wherein a is about 40 Å to about 60 Å, b is about 80 Å to about 100 Å, and c is about 40 Å to about 60 Å; and wherein $\alpha = \gamma = 90°$ and β is about 80° to about 120°. The crystallized enzyme is a dimer with a single dimer in the asymmetric unit.

Purified *S. aureus* thymidylate kinase at a concentration of about 1 mg/ml to about 50 mg/ml may be crystallized, for example, by using a streak seeding procedure from a solution including about 5 wt. % to about 50 wt. % PEG (preferably having a number average molecular weight between about 200 and about 20,000), about 0.05 M to about 0.5 M $MgCl_2$, and about 0 wt. % to about 20 wt. % DMSO, wherein the solution is buffered to a pH of about 6 to about 7. Use of a buffer having a pK_a of between about 5 and 8 is preferred. Molecular complexes of purified *S. aureus* thymidylate kinase at a concentration of about 1 mg/ml to about 50 mg/ml may also be crystallized, for example, from a solution including about 2 mM to about 20 mM β,γ-difluoromethylene-bisphosphonate adenosine monophosphate and about 0 wt. % to about 20 wt. % DMSO, wherein the solution is buffered to a pH of about 6 to about 7. A "molecular complex" means a protein in covalent or non-covalent association with a chemical entity. A buffer having a pK_a of between about 5 and 8 is preferred for use in the crystallization method. A particularly preferred buffer is about 0.4M to about 2.0M sodium citrate. Variation

in buffer and buffer pH as well as other additives such as PEG is apparent to those skilled in the art and may result in similar crystals.

The invention further includes an *S. aureus* thymidylate kinase crystal or *S. aureus* thymidylate kinase/ligand crystal that is isomorphous with an *S. aureus* thymidylate kinase crystal characterized by a unit cell having dimensions of a, b, and c; wherein a is about 40 Å to about 60 Å, b is about 80 Å to about 100 Å, and c is about 40 Å to about 60 Å; and wherein $\alpha = \gamma = 90^{\circ}$ and β is about 80° to about 120°.

X-ray Crystallographic Analysis

Each of the constituent amino acids of *S. aureus* thymidylate kinase is defined by a set of structure coordinates as set forth in Figure 2. The term "structure coordinates" refers to Cartesian coordinates derived from mathematical equations related to the patterns obtained on diffraction of a monochromatic beam of x-rays by the atoms (scattering centers) of an *S. aureus* thymidylate kinase complex in crystal form. The diffraction data are used to calculate an electron density map of the repeating unit of the crystal. The electron density maps are then used to establish the positions of the individual atoms of the *S. aureus* thymidylate kinase protein or protein/ligand complex.

Slight variations in structure coordinates can be generated by mathematically manipulating the *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase/ligand structure coordinates. For example, the structure coordinates set forth in Figure 2 could be manipulated by crystallographic permutations of the structure coordinates, fractionalization of the structure coordinates, integer additions or subtractions to sets of the structure coordinates, inversion of the structure coordinates or any combination of the above. Alternatively, modifications in the crystal structure due to mutations, additions, substitutions, and/or deletions of amino acids, or other changes in any of the components that make up the crystal, could also yield variations in structure coordinates. Such slight variations in the individual coordinates will have little

effect on overall shape. If such variations are within an acceptable standard error as compared to the original coordinates, the resulting three-dimensional shape is considered to be structurally equivalent. Structural equivalence is described in more detail below.

It should be noted that slight variations in individual structure coordinates of the *S. aureus* thymidylate kinase would not be expected to significantly alter the nature of chemical entities such as ligands that could associate with the substrate binding pockets. In this context, the phrase "associating with" refers to a condition of proximity between a chemical entity, or portions thereof, and an *S. aureus* thymidylate kinase molecule or portions thereof. The association may be non-covalent, wherein the juxtaposition is energetically favored by hydrogen bonding, van der Waals forces, or electrostatic interactions, or it may be covalent.

Thus, for example, a ligand that bound to a substrate binding pocket of *S. aureus* thymidylate kinase would also be expected to bind to or interfere with another substrate binding pocket whose structure coordinates define a shape that falls within the acceptable error.

It will be readily apparent to those of skill in the art that the numbering of amino acids in other isoforms of *S. aureus* thymidylate kinase may be different than that of *S. aureus* thymidylate kinase expressed in *E. coli*.

Active Site and Other Structural Features

Applicants' invention has provided, for the first time, information about the shape and structure of the substrate binding pockets of *S. aureus* thymidylate kinase. The structures of both the TMP and the TMP/ATP substrate binding pockets are elucidated. The secondary structure of the *S. aureus* thymidylate kinase monomer includes a five stranded parallel β sheet surrounded by nine α helices (Figure 3). This solved crystal structure of *S. aureus* thymidylate kinase does not contain any ligand which has resulted in a disordered loop between helices α7 and α8 (Figure 4). This loop has been called the "lid" in the structures of thymidylate kinase homologs from *E. coli* and *S. cerevisiae*. In *E. coli* the lid

contains Arg 153 which is responsible for phosphate binding of the ATP substrate as shown in the X-ray crystal structure of the *E. coli* enzyme with the AP_5T inhibitor, a transition state analog of TMP/ATP (A. Lavie et al., Biochemistry 37:3677-86 (1998)). In contrast the analogous arginine in *S. cerevisiae* comes from the P loop (Arg 15) between β1 and α1 (A. Lavie et al., Proc. Natl. Acad. Sci. USA, 95:14045-50 (1998)). This distinction as further manifested in sequence differences between the P loop and lid regions has led to the classification of the *S. cerevisiae* enzyme as a class I thymidylate kinase (which also includes human thymidylate kinase) and the *E. coli* enzyme as a class II thymidylate kinase (A. Lavie, Proc. Natl. Acad. Sci. USA, 95:14045-50 (1998)). Fortunately, *S. aureus* (SEQ ID NO:1) has greater sequence similarity to the *E. coli* enzyme (SEQ ID NO:2, 38% identical, 59% similar) than the *S. cerevisiae* enzyme (SEQ ID NO:3, 28% identical, 46% similar) and contains R148 in the lid region suggesting it should be classified as a class II thymidylate kinase (Figures 5 and 6). This classification suggests that it might be possible to design inhibitors that are specific for the *S. aureus* enzyme and not eukaryotic thymidylate kinases.

Superposition of the *S. aureus* TMK with *E. coli* TMK gave a r.m.s. deviation of 2.19 Å for analogous residues (Figure 5). Similarly, superposition of *S. aureus* TMK with *S. cerevisiae* TMK gave a r.m.s. deviation of 3.26 Å (Figure 6). Analysis of the active site residues from *E. coli* TMK as observed in the AP_5T inhibitor complex shows at least eleven residues that make direct hydrogen bonds to the inhibitor and another six residues make water mediated or hydrophobic interactions. Analysis of the active site residues from *S. aureus* TMK sequence reveals strong conservation of these active site residues with the *E.coli* active site (Figure 7b); 15 of the 17 residues involved in the protein-inhibitor complex are identical while the two remaining residues are strongly conserved. An analogous comparison for the *S. cerevisiae* TMK (Figure 7b) shows only four of 18 residues conserved within the active site suggesting that specificity between the *S. aureus* and eukaryotic thymidylate kinases might be attainable.

Comparing the liganded *E. coli* TMK structure with the unliganded *S. aureus* structure, it is apparent that a significant movement of the main chain around the active site (e.g. helix α2 and helix α7) including the ordering of the disordered residues must occur upon ligand binding. Figure 11 shows where the AP_5T inhibitor would be expected in the *S. aureus* TMK structure based on an alignment of the *E. coli* TMK-AP_5T inhibitor complex. There does appear to be a surface in the *S. aureus* TMK structure which would complement the ATP and TMP substrates, although it is clear from this surface view that an important part of the structure, the lid, is missing from the *S. aureus* TMK structure. Figure 8 clearly illustrates the role for this portion of the protein in completing the active site and closing off the thymidylate moiety from solvent.

Binding pockets are of significant utility in fields such as drug discovery. The association of natural ligands or substrates with the binding pockets of their corresponding receptors or enzymes is the basis of many biological mechanisms of action. Similarly, many drugs exert their biological effects through association with the binding pockets of receptors and enzymes. Such associations may occur with all or any parts of the binding pocket. An understanding of such associations helps lead to the design of drugs having more favorable associations with their target, and thus improved biological effects. Therefore, this information is valuable in designing potential inhibitors of *S. aureus* thymidylate kinase-like substrate binding pockets, as discussed in more detail below.

The term "binding pocket," as used herein, refers to a region of a molecule or molecular complex, that, as a result of its shape, favorably associates with another chemical entity. Thus, a binding pocket may include or consist of features such as cavities, surfaces, or interfaces between domains. Chemical entities that may associate with a binding pocket include, but are not limited to, cofactors, substrates, inhibitors, agonists, and antagonists.

The amino acid constituents of an *S. aureus* thymidylate kinase substrate binding pocket as defined herein are positioned in three dimensions in accordance with the structure coordinates listed in Figure 2. In one aspect, the

structure coordinates defining a substrate binding pocket of *S. aureus* thymidylate kinase include structure coordinates of all atoms in the constituent amino acids; in another aspect, the structure coordinates of a substrate binding pocket include structure coordinates of just the backbone atoms of the constituent atoms.

The TMP substrate binding pocket of *S. aureus* thymidylate kinase preferably includes the amino acids listed in Table 1, more preferably the amino acids listed in Table 2, and most preferably the amino acids listed in Table 3, as represented by the structure coordinates listed in Figure 2. Alternatively, the TMP substrate binding pocket of *S. aureus* thymidylate kinase may be defined by those amino acids whose backbone atoms are situated within about 3.5 Å, more preferably within about 5 Å, most preferably within about 7 Å, of one or more constituent atoms of a bound substrate or inhibitor. In yet another alternative, the TMP substrate binding pocket may be defined by those amino acids whose backbone atoms are situated within a sphere centered on the coordinates representing the alpha carbon atom of residue Ser98, the sphere having a radius of about 10 Å, preferably about 15 Å, and more preferably about 20 Å.

The TMP/ATP substrate binding pocket of *S. aureus* thymidylate kinase preferably includes the amino acids listed in Table 4, more preferably the amino acids listed in Table 5, and most preferably the amino acids listed in Table 6, as represented by the structure coordinates listed in Figure 2. Alternatively, the TMP/ATP substrate binding pocket of *S. aureus* thymidylate kinase may be defined by those amino acids whose backbone atoms are situated within about 3.5 Å, more preferably within about 5 Å, most preferably within about 7 Å, of one or more constituent atoms of a bound substrate or inhibitor. In yet another alternative, the TMP/ATP substrate binding pocket may be defined by those amino acids whose backbone atoms are situated within a sphere centered on the coordinates representing the alpha carbon atom of residue Arg93, the sphere having a radius of about 10 Å, preferably about 15 Å, and more preferably about 20 Å.

The term "*S. aureus* thymidylate kinase-like substrate binding pocket" refers to a portion of a molecule or molecular complex whose shape is

sufficiently similar to at least a portion of a substrate binding pocket of *S. aureus* thymidylate kinase as to be expected to bind related TMP and/or ATP structural analogues. A structurally equivalent substrate binding pocket is defined by a root mean square deviation from the structure coordinates of the backbone atoms of the amino acids that make up substrate binding pockets in *S. aureus* thymidylate kinase (as set forth in Figure 2) of at most about 2.1 Å. How this calculation is obtained is described below.

Accordingly, the invention provides molecules or molecular complexes comprising an *S. aureus* thymidylate kinase substrate binding pocket or *S. aureus* thymidylate kinase-like substrate binding pocket, as defined by the sets of structure coordinates described above.

Three-dimensional Configurations

X-ray structure coordinates define a unique configuration of points in space. Those of skill in the art understand that a set of structure coordinates for protein or an protein/ligand complex, or a portion thereof, define a relative set of points that, in turn, define a configuration in three dimensions. A similar or identical configuration can be defined by an entirely different set of coordinates, provided the distances and angles between coordinates remain essentially the same. In addition, a scalable configuration of points can be defined by increasing or decreasing the distances between coordinates by a scalar factor while keeping the angles essentially the same.

The present invention thus includes the scalable three-dimensional configuration of points derived from the structure coordinates of at least a portion of an *S. aureus* thymidylate kinase molecule or molecular complex, as listed in Figure 2, as well as structurally equivalent configurations, as described below. Preferably, the scalable three-dimensional configuration includes points derived from structure coordinates representing the locations of a plurality of the amino acids defining an *S. aureus* thymidylate kinase substrate binding pocket.

In one embodiment, the scalable three-dimensional configuration includes points derived from structure coordinates representing the locations the backbone atoms of a plurality of amino acids defining the *S. aureus* thymidylate kinase TMP substrate binding pocket, preferably the amino acids listed in Table 1, more preferably the amino acids listed in Table 2, and most preferably the amino acids listed in Table 3. Alternatively, the scalable three-dimensional configuration includes points derived from structure coordinates representing the locations of the side chain and the backbone atoms (other than hydrogens) of a plurality of the amino acids defining the *S. aureus* thymidylate kinase TMP substrate binding pocket, preferably the amino acids listed in Table 1, more preferably the amino acids listed in Table 2, and most preferably the amino acids listed in Table 3.

In another embodiment, the scalable three-dimensional configuration includes points derived from structure coordinates representing the locations the backbone atoms of a plurality of amino acids defining the *S. aureus* thymidylate kinase TMP/ATP substrate binding pocket, preferably the amino acids listed in Table 4, more preferably the amino acids listed in Table 5, and most preferably the amino acids listed in Table 6. Alternatively, the scalable three-dimensional configuration includes points derived from structure coordinates representing the locations of the side chain and the backbone atoms (other than hydrogens) of a plurality of the amino acids defining the *S. aureus* thymidylate kinase TMP/ATP substrate binding pocket, preferably the amino acids listed in Table 4, more preferably the amino acids listed in Table 5, and most preferably the amino acids listed in Table 6.

Likewise, the invention also includes the scalable three-dimensional configuration of points derived from structure coordinates of molecules or molecular complexes that are structurally homologous to *S. aureus* thymidylate kinase, as well as structurally equivalent configurations. Structurally homologous molecules or molecular complexes are defined below. Advantageously, structurally homologous molecules can be identified using the structure coordinates of *S. aureus* thymidylate kinase according to a method of the invention.

The configurations of points in space derived from structure coordinates according to the invention can be visualized as, for example, a holographic image, a stereodiagram, a model or a computer-displayed image, and the invention thus includes such images, diagrams or models.

Structurally Equivalent Crystal Structures

Various computational analyses can be used to determine whether a molecule or a substrate binding pocket portion thereof is "structurally equivalent," defined in terms of its three-dimensional structure, to all or part of *S. aureus* thymidylate kinase or its substrate binding pockets. Such analyses may be carried out in current software applications, such as the Molecular Similarity application of QUANTA (Molecular Simulations Inc., San Diego, CA) version 4.1, and as described in the accompanying User's Guide.

The Molecular Similarity application permits comparisons between different structures, different conformations of the same structure, and different parts of the same structure. The procedure used in Molecular Similarity to compare structures is divided into four steps: (1) load the structures to be compared; (2) define the atom equivalences in these structures; (3) perform a fitting operation; and (4) analyze the results.

Each structure is identified by a name. One structure is identified as the target (i.e., the fixed structure); all remaining structures are working structures (i.e., moving structures). Since atom equivalency within QUANTA is defined by user input, for the purpose of this invention equivalent atoms are defined as protein backbone atoms (N, Cα, C, and O) for all conserved residues between the two structures being compared. A conserved residue is defined as a residue which is structurally or functionally equivalent. Only rigid fitting operations are considered.

When a rigid fitting method is used, the working structure is translated and rotated to obtain an optimum fit with the target structure. The fitting operation uses an algorithm that computes the optimum translation and rotation to be applied to the moving structure, such that the root mean square difference of the

fit over the specified pairs of equivalent atom is an absolute minimum. This number, given in angstroms, is reported by QUANTA.

For the purpose of this invention, any molecule or molecular complex or substrate binding pocket thereof, or any portion thereof, that has a root mean square deviation of conserved residue backbone atoms (N, Cα, C, O) of less than about 2.1 Å, when superimposed on the relevant backbone atoms described by the reference structure coordinates listed in Figure 2, is considered "structurally equivalent" to the reference molecule. That is to say, the crystal structures of those portions of the two molecules are substantially identical, within acceptable error. Particularly preferred structurally equivalent molecules or molecular complexes are those that are defined by the entire set of structure coordinates listed in Figure 2 ± a root mean square deviation from the conserved backbone atoms of those amino acids of not more than 2.1 Å. More preferably, the root mean square deviation is less than about 1.0 Å. Another embodiment of this invention is a molecular complex defined by the structure coordinates listed in Figure 2 for those amino acids listed in Table 1, ± a root mean square deviation from the conserved backbone atoms of those amino acids of not more than 2.1 Å, preferably less than about 1.0 Å. Still another embodiment of this invention is a molecular complex defined by the structure coordinates listed in Figure 2 for those amino acids listed in Table 4, ± a root mean square deviation from the conserved backbone atoms of those amino acids of not more than 2.1 Å, preferably less than about 1.0 Å.

The term "root mean square deviation" means the square root of the arithmetic mean of the squares of the deviations. It is a way to express the deviation or variation from a trend or object. For purposes of this invention, the "root mean square deviation" defines the variation in the backbone of a protein from the backbone of *S. aureus* thymidylate kinase or a substrate binding pocket portion thereof, as defined by the structure coordinates of *S. aureus* thymidylate kinase described herein.

Machine Readable Storage Media

Transformation of the structure coordinates for all or a portion of *S. aureus* thymidylate kinase or the *S. aureus* thymidylate kinase/ligand complex or one of its substrate binding pockets, for structurally homologous molecules as defined below, or for the structural equivalents of any of these molecules or molecular complexes as defined above, into three-dimensional graphical representations of the molecule or complex can be conveniently achieved through the use of commercially-available software.

The invention thus further provides a machine-readable storage medium comprising a data storage material encoded with machine readable data which, when using a machine programmed with instructions for using said data, is capable of displaying a graphical three-dimensional representation of any of the molecule or molecular complexes of this invention that have been described above. In a preferred embodiment, the machine-readable data storage medium comprises a data storage material encoded with machine readable data which, when using a machine programmed with instructions for using said data, is capable of displaying a graphical three-dimensional representation of a molecule or molecular complex comprising all or any parts of an *S. aureus* thymidylate kinase substrate binding pocket or an *S. aureus* thymidylate kinase-like substrate binding pocket, as defined above. In another preferred embodiment, the machine-readable data storage medium comprises a data storage material encoded with machine readable data which, when using a machine programmed with instructions for using said data, is capable of displaying a graphical three-dimensional representation of a molecule or molecular complex defined by the structure coordinates of all of the amino acids listed in Figure 2, ± a root mean square deviation from the backbone atoms of said amino acids of not more than 2.1 Å.

In an alternative embodiment, the machine-readable data storage medium comprises a data storage material encoded with a first set of machine readable data which comprises the Fourier transform of the structure coordinates set forth in Figure 2, and which, when using a machine programmed with instructions

for using said data, can be combined with a second set of machine readable data comprising the x-ray diffraction pattern of a molecule or molecular complex to determine at least a portion of the structure coordinates corresponding to the second set of machine readable data.

For example, a system for reading a data storage medium may include a computer comprising a central processing unit ("CPU"), a working memory which may be, e.g., RAM (random access memory) or "core" memory, mass storage memory (such as one or more disk drives or CD-ROM drives), one or more display devices (e.g., cathode-ray tube ("CRT") displays, light emitting diode ("LED") displays, liquid crystal displays ("LCDs"), electroluminescent displays, vacuum fluorescent displays, field emission displays ("FEDs"), plasma displays, projection panels, etc.), one or more user input devices (e.g., keyboards, microphones, mice, track balls, touch pads, etc.), one or more input lines, and one or more output lines, all of which are interconnected by a conventional bidirectional system bus. The system may be a stand-alone computer, or may be networked (e.g., through local area networks, wide area networks, intranets, extranets, or the internet) to other systems (e.g., computers, hosts, servers, etc.). The system may also include additional computer controlled devices such as consumer electronics and appliances.

Input hardware may be coupled to the computer by input lines and may be implemented in a variety of ways. Machine-readable data of this invention may be inputted via the use of a modem or modems connected by a telephone line or dedicated data line. Alternatively or additionally, the input hardware may comprise CD-ROM drives or disk drives. In conjunction with a display terminal, a keyboard may also be used as an input device.

Output hardware may be coupled to the computer by output lines and may similarly be implemented by conventional devices. By way of example, the output hardware may include a display device for displaying a graphical representation of a binding pocket of this invention using a program such as QUANTA as described herein. Output hardware might also include a printer, so

that hard copy output may be produced, or a disk drive, to store system output for later use.

In operation, a CPU coordinates the use of the various input and output devices, coordinates data accesses from mass storage devices, accesses to and from working memory, and determines the sequence of data processing steps. A number of programs may be used to process the machine-readable data of this invention. Such programs are discussed in reference to the computational methods of drug discovery as described herein. References to components of the hardware system are included as appropriate throughout the following description of the data storage medium.

Machine-readable storage devices useful in the present invention include, but are not limited to, magnetic devices, electrical devices, optical devices, and combinations thereof. Examples of such data storage devices include, but are not limited to, hard disk devices, CD devices, digital video disk devices, floppy disk devices, removable hard disk devices, magneto-optic disk devices, magnetic tape devices, flash memory devices, bubble memory devices, holographic storage devices, and any other mass storage peripheral device. It should be understood that these storage devices include necessary hardware (e.g., drives, controllers, power supplies, etc.) as well as any necessary media (e.g., disks, flash cards, etc.) to enable the storage of data.

Structurally Homologous Molecules, Molecular Complexes, and Crystal Structures

The structure coordinates set forth in Figure 2 can be used to aid in obtaining structural information about another crystallized molecule or molecular complex. The method of the invention allows determination of at least a portion of the three-dimensional structure of molecules or molecular complexes which contain one or more structural features that are similar to structural features of *S. aureus* thymidylate kinase, These molecules are referred to herein as "structurally homologous" to *S. aureus* thymidylate kinase, Similar structural features can

include, for example, regions of amino acid identity, conserved active site or binding site motifs, and similarly arranged secondary structural elements (e.g., α helices and β sheets). Optionally, structural homology is determined by aligning the residues of the two amino acid sequences to optimize the number of identical amino acids along the lengths of their sequences; gaps in either or both sequences are permitted in making the alignment in order to optimize the number of identical amino acids, although the amino acids in each sequence must nonetheless remain in their proper order. Preferably, two amino acid sequences are compared using the Blastp program, version 2.0.9, of the BLAST 2 search algorithm, as described by Tatiana et al., FEMS Microbiol Lett 174, 247-50 (1999), and available at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. Preferably, the default values for all BLAST 2 search parameters are used, including matrix = BLOSUM62; open gap penalty = 11, extension gap penalty = 1, gap x_dropoff = 50, expect = 10, wordsize = 3, and filter on. In the comparison of two amino acid sequences using the BLAST search algorithm, structural similarity is referred to as "identity." Preferably, a structurally homologous molecule is a protein that has an amino acid sequence sharing at least 65% identity with a native or recombinant amino acid sequence of *S. aureus* thymidylate kinase (for example, SEQ ID NO:1). More preferably, a protein that is structurally homologous to *S. aureus* thymidylate kinase includes at least one contiguous stretch of at least 50 amino acids that shares at least 80% amino acid sequence identity with the analogous portion of the native or recombinant *S. aureus* thymidylate kinase (for example, SEQ ID NO:1). Methods for generating structural information about the structurally homologous molecule or molecular complex are well-known and include, for example, molecular replacement techniques.

Therefore, in another embodiment this invention provides a method of utilizing molecular replacement to obtain structural information about a molecule or molecular complex whose structure is unknown comprising the steps of:

(a) crystallizing the molecule or molecular complex of unknown structure;

(b) generating an x-ray diffraction pattern from said crystallized molecule or molecular complex; and

(c) applying at least a portion of the structure coordinates set forth in Figure 2 to the x-ray diffraction pattern to generate a three-dimensional electron density map of the molecule or molecular complex whose structure is unknown.

By using molecular replacement, all or part of the structure coordinates of *S. aureus* thymidylate kinase or the *S. aureus* thymidylate kinase/ligand complex as provided by this invention can be used to determine the structure of a crystallized molecule or molecular complex whose structure is unknown more quickly and efficiently than attempting to determine such information *ab initio*.

Molecular replacement provides an accurate estimation of the phases for an unknown structure. Phases are a factor in equations used to solve crystal structures that cannot be determined directly. Obtaining accurate values for the phases, by methods other than molecular replacement, is a time-consuming process that involves iterative cycles of approximations and refinements and greatly hinders the solution of crystal structures. However, when the crystal structure of a protein containing at least a structurally homologous portion has been solved, the phases from the known structure provide a satisfactory estimate of the phases for the unknown structure.

Thus, this method involves generating a preliminary model of a molecule or molecular complex whose structure coordinates are unknown, by orienting and positioning the relevant portion of *S. aureus* thymidylate kinase or the *S. aureus* thymidylate kinase/ligand complex according to Figure 2 within the unit cell of the crystal of the unknown molecule or molecular complex so as best to account for the observed x-ray diffraction pattern of the crystal of the molecule or molecular complex whose structure is unknown. Phases can then be calculated from this model and combined with the observed x-ray diffraction pattern amplitudes to

generate an electron density map of the structure whose coordinates are unknown. This, in turn, can be subjected to any well-known model building and structure refinement techniques to provide a final, accurate structure of the unknown crystallized molecule or molecular complex (E. Lattman, "Use of the Rotation and Translation Functions," in Meth. Enzymol., 115, pp. 55-77 (1985); M.G. Rossman, ed., "The Molecular Replacement Method," Int. Sci. Rev. Ser., No. 13, Gordon & Breach, New York (1972)).

Structural information about a portion of any crystallized molecule or molecular complex that is sufficiently structurally homologous to a portion of *S. aureus* thymidylate kinase can be resolved by this method. In addition to a molecule that shares one or more structural features with *S. aureus* thymidylate kinase as described above, a molecule that has similar bioactivity, such as the same catalytic activity, substrate specificity or ligand binding activity as *S. aureus* thymidylate kinase, may also be sufficiently structurally homologous to *S. aureus* thymidylate kinase to permit use of the structure coordinates of *S. aureus* thymidylate kinase to solve its crystal structure.

In a preferred embodiment, the method of molecular replacement is utilized to obtain structural information about a molecule or molecular complex, wherein the molecule or molecular complex comprises at least one *S. aureus* thymidylate kinase subunit or homolog. A "subunit" of *S. aureus* thymidylate kinase is an *S. aureus* thymidylate kinase molecule that has been truncated at the N-terminus or the C-terminus, or both. In the context of the present invention, a "homolog" of *S. aureus* thymidylate kinase is a protein that contains one or more amino acid substitutions, deletions, additions, or rearrangements with respect to the amino acid sequence of *S. aureus* thymidylate kinase (SEQ ID NO:1), but that, when folded into its native conformation, exhibits or is reasonably expected to exhibit at least a portion of the tertiary (three-dimensional) structure of *S. aureus* thymidylate kinase, For example, structurally homologous molecules can contain deletions or additions of one or more contiguous or noncontiguous amino acids, such as a loop or a domain. Structurally homologous molecules also include

"modified" *S. aureus* thymidylate kinase molecules that have been chemically or enzymatically derivatized at one or more constituent amino acid, including side chain modifications, backbone modifications, and N- and C- terminal modifications including acetylation, hydroxylation, methylation, amidation, and the attachment of carbohydrate or lipid moieties, cofactors, and the like.

A heavy atom derivative of *S. aureus* thymidylate kinase is also included as an *S. aureus* thymidylate kinase homolog. The term "heavy atom derivative" refers to derivatives of *S. aureus* thymidylate kinase produced by chemically modifying a crystal of *S. aureus* thymidylate kinase, In practice, a crystal is soaked in a solution containing heavy metal atom salts, or organometallic compounds, e.g., lead chloride, gold thiomalate, thiomersal or uranyl acetate, which can diffuse through the crystal and bind to the surface of the protein. The location(s) of the bound heavy metal atom(s) can be determined by x-ray diffraction analysis of the soaked crystal. This information, in turn, is used to generate the phase information used to construct three-dimensional structure of the protein (T.L. Blundell and N.L. Johnson, Protein Crystallography, Academic Press (1976)).

Because *S. aureus* thymidylate kinase can crystallize in more than one crystal form, the structure coordinates of *S. aureus* thymidylate kinase as provided by this invention are particularly useful in solving the structure of other crystal forms of *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase complexes.

The structure coordinates of *S. aureus* thymidylate kinase as provided by this invention are particularly useful in solving the structure of *S. aureus* thymidylate kinase mutants. Mutants may be prepared, for example, by expression of *S. aureus* thymidylate kinase cDNA previously altered in its coding sequence by oligonucleotide-directed mutagenesis. Mutants may also be generated by site-specific incorporation of unnatural amino acids into thymidylate kinase proteins using the general biosynthetic method of C.J. Noren et al., Science, 244:182-188 (1989). In this method, the codon encoding the amino acid of interest in wild-type *S. aureus* thymidylate kinase is replaced by a "blank" nonsense codon, TAG, using

oligonucleotide-directed mutagenesis. A suppressor tRNA directed against this codon is then chemically aminoacylated *in vitro* with the desired unnatural amino acid. The aminoacylated tRNA is then added to an *in vitro* translation system to yield a mutant *S. aureus* thymidylate kinase with the site-specific incorporated unnatural amino acid.

Selenocysteine or selenomethionine may be incorporated into wild-type or mutant *S. aureus* thymidylate kinase by expression of *S. aureus* thymidylate kinase-encoding cDNAs in auxotrophic *E. coli* strains (W.A. Hendrickson et al., EMBO J., 9(5):1665-1672 (1990)). In this method, the wild-type or mutagenized *S. aureus* thymidylate kinase cDNA may be expressed in a host organism on a growth medium depleted of either natural cysteine or methionine (or both) but enriched in selenocysteine or selenomethionine (or both). Alternatively, selenomethionine analogues may be prepared by down regulation methionine biosynthesis. (T.E. Benson et al., Nat. Struct. Biol., 2:644-53 (1995); G.D. Van Duyne et al., J. Mol. Biol. 229:105-24 (1993)).

The structure coordinates of *S. aureus* thymidylate kinase listed in Figure 2 are also particularly useful to solve the structure of crystals of *S. aureus* thymidylate kinase, *S. aureus* thymidylate kinase mutants or *S. aureus* thymidylate kinase homologs co-complexed with a variety of chemical entities. This approach enables the determination of the optimal sites for interaction between chemical entities, including candidate *S. aureus* thymidylate kinase inhibitors and *S. aureus* thymidylate kinase. Potential sites for modification within the various binding site of the molecule can also be identified. This information provides an additional tool for determining the most efficient binding interactions, for example, increased hydrophobic interactions, between *S. aureus* thymidylate kinase and a chemical entity. For example, high resolution x-ray diffraction data collected from crystals exposed to different types of solvent allows the determination of where each type of solvent molecule resides. Small molecules that bind tightly to those sites can then be designed and synthesized and tested for their *S. aureus* thymidylate kinase inhibition activity.

All of the complexes referred to above may be studied using well-known x-ray diffraction techniques and may be refined versus 1.5-3 Å resolution x-ray data to an R value of about 0.20 or less using computer software, such as X-PLOR (Yale University, 81992, distributed by Molecular Simulations, Inc.; see, e.g., Blundell & Johnson, supra; Meth. Enzymol., Vol. 114 & 115, H.W. Wyckoff et al., eds., Academic Press (1985)). This information may thus be used to optimize known *S. aureus* thymidylate kinase inhibitors, and more importantly, to design new *S. aureus* thymidylate kinase inhibitors.

The invention also includes the unique three-dimensional configuration defined by a set of points defined by the structure coordinates for a molecule or molecular complex structurally homologous to *S. aureus* thymidylate kinase as determined using the method of the present invention, structurally equivalent configurations, and magnetic storage media comprising such set of structure coordinates.

Further, the invention includes structurally homologous molecules as identified using the method of the invention.

Homology Modeling

Using homology modeling, a computer model of an *S. aureus* thymidylate kinase homolog can be built or refined without crystallizing the homolog. First, a preliminary model of the *S. aureus* thymidylate kinase homolog is created by sequence alignment with *S. aureus* thymidylate kinase, secondary structure prediction, the screening of structural libraries, or any combination of those techniques. Computational software may be used to carry out the sequence alignments and the secondary structure predictions. Structural incoherences, e.g., structural fragments around insertions and deletions, can be modeled by screening a structural library for peptides of the desired length and with a suitable conformation. For prediction of the side chain conformation, a side chain rotamer library may be employed. If the *S. aureus* thymidylate kinase homolog has been crystallized, the

final homology model can be used to solve the crystal structure of the homolog by molecular replacement, as described above. Next, the preliminary model is subjected to energy minimization to yield an energy minimized model. The energy minimized model may contain regions where stereochemistry restraints are violated, in which case such regions are remodeled to obtain a final homology model. The homology model is positioned according to the results of molecular replacement, and subjected to further refinement comprising molecular dynamics calculations.

Rational Drug Design

Computational techniques can be used to screen, identify, select and/or design chemical entities capable of associating with *S. aureus* thymidylate kinase or structurally homologous molecules. Knowledge of the structure coordinates for *S. aureus* thymidylate kinase permits the design and/or identification of synthetic compounds and/or other molecules which have a shape complementary to the conformation of the *S. aureus* thymidylate kinase binding site. In particular, computational techniques can be used to identify or design chemical entities, such as inhibitors, agonists and antagonists, that associate with an *S. aureus* thymidylate kinase substrate binding pocket or an *S. aureus* thymidylate kinase-like substrate binding pocket. Inhibitors may bind to or interfere with all or a portion of an active site of *S. aureus* thymidylate kinase, and can be competitive, non-competitive, or uncompetitive inhibitors; or interfere with dimerization by binding at the interface between the two monomers. Once identified and screened for biological activity, these inhibitors/agonists/antagonists may be used therapeutically or prophylactically to block *S. aureus* thymidylate kinase activity and, thus, inhibit the growth of the bacteria or cause its death. Structure-activity data for analogues of ligands that bind to or interfere with *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pockets can also be obtained computationally.

The term "chemical entity," as used herein, refers to chemical compounds, complexes of two or more chemical compounds, and fragments of such

compounds or complexes. Chemical entities that are determined to associate with *S. aureus* thymidylate kinase are potential drug candidates.

Data stored in a machine-readable storage medium that is capable of displaying a graphical three-dimensional representation of the structure of *S. aureus* thymidylate kinase or a structurally homologous molecule, as identified herein, or portions thereof may thus be advantageously used for drug discovery. The structure coordinates of the chemical entity are used to generate a three-dimensional image that can be computationally fit to the three-dimensional image of *S. aureus* thymidylate kinase or a structurally homologous molecule. The three-dimensional molecular structure encoded by the data in the data storage medium can then be computationally evaluated for its ability to associate with chemical entities. When the molecular structures encoded by the data is displayed in a graphical three-dimensional representation on a computer screen, the protein structure can also be visually inspected for potential association with chemical entities.

One embodiment of the method of drug design involves evaluating the potential association of a known chemical entity with *S. aureus* thymidylate kinase or a structurally homologous molecule, particularly with an *S. aureus* thymidylate kinase substrate binding pocket or *S. aureus* thymidylate kinase-like substrate binding pocket. The method of drug design thus includes computationally evaluating the potential of a selected chemical entity to associate with any of the molecules or molecular complexes set forth above. This method comprises the steps of: (a) employing computational means to perform a fitting operation between the selected chemical entity and a substrate binding pocket or a pocket nearby the substrate binding pocket of the molecule or molecular complex; and (b) analyzing the results of said fitting operation to quantify the association between the chemical entity and the substrate binding pocket.

In another embodiment, the method of drug design involves computer-assisted design of chemical entities that associate with *S. aureus* thymidylate kinase, its homologs, or portions thereof. Chemical entities can be designed in a step-wise fashion, one fragment at a time, or may be designed as a whole or "*de novo*."

To be a viable drug candidate, the chemical entity identified or designed according to the method must be capable of structurally associating with at least part of an *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pockets, and must be able, sterically and energetically, to assume a conformation that allows it to associate with the *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket. Non-covalent molecular interactions important in this association include hydrogen bonding, van der Waals interactions, hydrophobic interactions, and electrostatic interactions. Conformational considerations include the overall three-dimensional structure and orientation of the chemical entity in relation to the substrate binding pocket, and the spacing between various functional groups of an entity that directly interact with the *S. aureus* thymidylate kinase-like substrate binding pocket or homologs thereof.

Optionally, the potential binding of a chemical entity to an *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket is analyzed using computer modeling techniques prior to the actual synthesis and testing of the chemical entity. If these computational experiments suggest insufficient interaction and association between it and the *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket, testing of the entity is obviated. However, if computer modeling indicates a strong interaction, the molecule may then be synthesized and tested for its ability to bind to or interfere with an *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket. Binding assays to determine if a compound actually interferes with *S. aureus* thymidylate kinase can also be performed and are well known in the art. Binding assays may employ kinetic or thermodynamic methodology using a wide variety of techniques including, but not limited to, microcalorimetry, circular

dichroism, capillary zone electrophoresis, nuclear magnetic resonance spectroscopy, fluorescence spectroscopy, and combinations thereof.

One skilled in the art may use one of several methods to screen chemical entities or fragments for their ability to associate with an *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket. This process may begin by visual inspection of, for example, an *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket on the computer screen based on the *S. aureus* thymidylate kinase structure coordinates listed in Figure 2 or other coordinates which define a similar shape generated from the machine-readable storage medium. Selected fragments or chemical entities may then be positioned in a variety of orientations, or docked, within the substrate binding pocket. Docking may be accomplished using software such as QUANTA and SYBYL, followed by energy minimization and molecular dynamics with standard molecular mechanics forcefields, such as CHARMM and AMBER.

Specialized computer programs may also assist in the process of selecting fragments or chemical entities. Examples include GRID (P.J. Goodford, J. Med. Chem. 28:849-857 (1985); available from Oxford University, Oxford, UK); MCSS (A. Miranker et al., Proteins: Struct. Funct. Gen.,11:29-34 (1991); available from Molecular Simulations, San Diego, CA); AUTODOCK (D.S. Goodsell et al., Proteins: Struct. Funct. Genet. 8:195-202 (1990); available from Scripps Research Institute, La Jolla, CA); and DOCK (I.D. Kuntz et al., J. Mol. Biol. 161:269-288 (1982); available from University of California, San Francisco, CA).

Once suitable chemical entities or fragments have been selected, they can be assembled into a single compound or complex. Assembly may be preceded by visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of *S. aureus* thymidylate kinase. This would be followed by manual model building using software such as QUANTA or SYBYL (Tripos Associates, St. Louis, MO).

Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include, without limitation, CAVEAT (P.A. Bartlett et al., in Molecular Recognition in Chemical and Biological Problems," Special Publ., Royal Chem. Soc., 78:182-196 (1989); G. Lauri et al., J. Comput. Aided Mol. Des. 8:51-66 (1994); available from the University of California, Berkeley, CA); 3D database systems such as ISIS (available from MDL Information Systems, San Leandro, CA; reviewed in Y.C. Martin, J. Med. Chem. 35:2145-2154 (1992)); and HOOK (M.B. Eisen et al., Proteins: Struc., Funct., Genet. 19:199-221 (1994); available from Molecular Simulations, San Diego, CA).

S. aureus thymidylate kinase binding compounds may be designed "*de novo*" using either an empty binding site or optionally including some portion(s) of a known inhibitor(s). There are many *de novo* ligand design methods including, without limitation, LUDI (H.-J. Bohm, J. Comp. Aid. Molec. Design. 6:61-78 (1992); available from Molecular Simulations Inc., San Diego, CA); LEGEND (Y. Nishibata et al., Tetrahedron, 47:8985 (1991); available from Molecular Simulations Inc., San Diego, CA); LeapFrog (available from Tripos Associates, St. Louis, MO); and SPROUT (V. Gillet et al., J. Comput. Aided Mol. Design 7:127-153 (1993); available from the University of Leeds, UK).

Once a compound has been designed or selected by the above methods, the efficiency with which that entity may bind to or interfere with an *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket may be tested and optimized by computational evaluation. For example, an effective *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket inhibitor must preferably demonstrate a relatively small difference in energy between its bound and free states (i.e., a small deformation energy of binding). Thus, the most efficient *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket inhibitors should preferably be designed with a deformation energy of binding of not greater than about 10 kcal/mole; more preferably, not greater than 7 kcal/mole. *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket inhibitors may

interact with the substrate binding pocket in more than one conformation that is similar in overall binding energy. In those cases, the deformation energy of binding is taken to be the difference between the energy of the free entity and the average energy of the conformations observed when the inhibitor binds to the protein.

An entity designed or selected as binding to or interfering with an *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket may be further computationally optimized so that in its bound state it would preferably lack repulsive electrostatic interaction with the target enzyme and with the surrounding water molecules. Such non-complementary electrostatic interactions include repulsive charge-charge, dipole-dipole, and charge-dipole interactions.

Specific computer software is available in the art to evaluate compound deformation energy and electrostatic interactions. Examples of programs designed for such uses include: Gaussian 94, revision C (M.J. Frisch, Gaussian, Inc., Pittsburgh, PA 81995); AMBER, version 4.1 (P.A. Kollman, University of California at San Francisco, 81995); QUANTA/CHARMM (Molecular Simulations, Inc., San Diego, CA 81995); Insight II/Discover (Molecular Simulations, Inc., San Diego, CA 81995); DelPhi (Molecular Simulations, Inc., San Diego, CA 81995); and AMSOL (Quantum Chemistry Program Exchange, Indiana University). These programs may be implemented, for instance, using a Silicon Graphics workstation such as an Indigo^2 with "IMPACT" graphics. Other hardware systems and software packages will be known to those skilled in the art.

Another approach encompassed by this invention is the computational screening of small molecule databases for chemical entities or compounds that can bind in whole, or in part, to a *S. aureus* thymidylate kinase or *S. aureus* thymidylate kinase-like substrate binding pocket. In this screening, the quality of fit of such entities to the binding site may be judged either by shape complementarity or by estimated interaction energy (E.C. Meng et al., J. Comp. Chem., 13, pp. 505-524 (1992)).

This invention also enables the development of chemical entities that can isomerize to short-lived reaction intermediates in the chemical reaction of a substrate or other compound that interferes with or with *S. aureus* thymidylate kinase. Time-dependent analysis of structural changes in *S. aureus* thymidylate kinase during its interaction with other molecules is carried out. The reaction intermediates of *S. aureus* thymidylate kinase can also be deduced from the reaction product in co-complex with *S. aureus* thymidylate kinase. Such information is useful to design improved analogues of known *S. aureus* thymidylate kinase inhibitors or to design novel classes of inhibitors based on the reaction intermediates of the *S. aureus* thymidylate kinase and inhibitor co-complex. This provides a novel route for designing *S. aureus* thymidylate kinase inhibitors with both high specificity and stability.

Yet another approach to rational drug design involves probing the *S. aureus* thymidylate kinase crystal of the invention with molecules comprising a variety of different functional groups to determine optimal sites for interaction between candidate *S. aureus* thymidylate kinase inhibitors and the protein. For example, high resolution x-ray diffraction data collected from crystals soaked in or co-crystallized with other molecules allows the determination of where each type of solvent molecule sticks. Molecules that bind tightly to those sites can then be further modified and synthesized and tested for their thymidylate kinase inhibitor activity (J. Travis, Science, 262:1374 (1993)).

In a related approach, iterative drug design is used to identify inhibitors of *S. aureus* thymidylate kinase. Iterative drug design is a method for optimizing associations between a protein and a compound by determining and evaluating the three-dimensional structures of successive sets of protein/compound complexes. In iterative drug design, crystals of a series of protein/compound complexes are obtained and then the three-dimensional structures of each complex is solved. Such an approach provides insight into the association between the proteins and compounds of each complex. This is accomplished by selecting compounds with inhibitory activity, obtaining crystals of this new protein/compound complex,

solving the three dimensional structure of the complex, and comparing the associations between the new protein/compound complex and previously solved protein/compound complexes. By observing how changes in the compound affected the protein/compound associations, these associations may be optimized.

A compound that is identified or designed as a result of any of these methods can be obtained (or synthesized) and tested for its biological activity, e.g., inhibition of thymidylate kinase activity.

Pharmaceutical Compositions (Inhibitors)

Pharmaceutical compositions of this invention comprise an inhibitor of *S. aureus* TMK activity identified according to the invention, or a pharmaceutically acceptable salt thereof, and a pharmaceutically acceptable carrier, adjuvant, or vehicle. The term "pharmaceutically acceptable carrier" refers to a carrier(s) that is "acceptable" in the sense of being compatible with the other ingredients of a composition and not deleterious to the recipient thereof. Optionally, the pH of the formulation is adjusted with pharmaceutically acceptable acids, bases, or buffers to enhance the stability of the formulated compound or its delivery form.

Methods of making and using such pharmaceutical compositions are also included in the invention. The pharmaceutical compositions of the invention can be administered orally, parenterally, by inhalation spray, topically, rectally, nasally, buccally, vaginally, or via an implanted reservoir. Oral administration or administration by injection is preferred. The term parenteral as used herein includes subcutaneous, intracutaneous, intravenous, intramuscular, intra-articular, intrasynovial, intrasternal, intrathecal, intralesional, and intracranial injection or infusion techniques.

Dosage levels of between about 0.01 and about 100 mg/kg body weight per day, preferably between about 0.5 and about 75 mg/kg body weight per day of the *S. aureus* TMK inhibitory compounds described herein are useful for the prevention and treatment of *S. aureus* TMK mediated disease. Typically, the pharmaceutical compositions of this invention will be administered from about 1 to

about 5 times per day or alternatively, as a continuous infusion. Such administration can be used as a chronic or acute therapy. The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. A typical preparation will contain from about 5% to about 95% active compound (w/w). Preferably, such preparations contain from about 20% to about 80% active compound.

In order that this invention be more fully understood, the following examples are set forth. These examples are for the purpose of illustration only and are not to be construed as limiting the scope of the invention in any way.

EXAMPLES

Example 1: Analysis of the Structure of *S. aureus* Thymidylate Kinase

A. Expression, Purification and Crystallization

The M15-1C *Escherichia coli* construct expressing *S. aureus* thymidylate kinase was obtained as a strain in which the Qiagen pREP4 vector was replaced with pREP4UX. Genes and polypeptides derived from *S. aureus*, including *S. aureus* and thymidylate kinase, are published in EP 786519 A2 and WO 0012678, both assigned to Human Genome Sciences. This plasmid contains the argU gene which codes for the AGA tRNA and prevents the lysine for arginine substitution which occurred in the original construct from Human Genome Sciences. For preparation of the selenomethionine analogue of thymidylate kinase, the construct was grown in a minimal salts medium, M9, which contained glucose and NH_4Cl as the sources of carbon and nitrogen. Endogenous methionine biosynthesis was then inhibited while adding an excess of selenomethionine to the growth medium just prior to IPTG induction of thymidylate kinase synthesis (T.E. Benson et al., Nat. Struct. Biol., 2:644-53 (1995); G.D. Van Duyne et al., J. Mol. Biol. 229:105-24 (1993)). The formulation of basal M9 was Na_2HPO_4, 6 g; KH_2PO_4, 3 g; NH_4Cl, 1.0

g; and NaCl, 0.5 g per L of deionized water. The pH was adjusted to 7.4 with concentrated KOH and the medium was sterilized by autoclaving. Prior to inoculation, the following filter sterilized solutions were added per L of basal medium: 1M $MgSO_4$, 1.0 mL; 1M $CaCl_2$, 0.1 mL; trace metal salts solution, 0.1 mL, 10 mM thiamin, 1.0 mL; and 20% glucose, 20 mL. The trace metal salts solution contained per L of deionized water: $MgCl_2 \cdot 6H_2O$, 39.44 g; $MnSO_4 \cdot H_2O$, 5.58 g; $FeSO_4 \cdot 7H_2O$, 1.11 g; $Na_2MoO_4 \cdot 2H_2O$, 0.48 g; $CaCl_2$, 0.33 g; NaCl, 0.12 g; and ascorbic acid, 1.0 g. Filter sterilized ampicillin and kanamycin were added to the medium at final concentrations of 100 mg/mL and 30mg/mL, respectively.

Fermentations were prepared in 100 mL volumes of M9 medium contained in 500 mL wide mouth flasks. A 0.1 mL aliquot of the stock culture was inoculated into the medium and allowed to grow at 30°C for 18 - 20 hours with a shaking rate of 200 rpm. The seed culture was harvested by centrifugation and then resuspended in an equal volume of M9 medium. The resuspended seed was used to inoculate expression fermentations at a rate of 3%. For expression, the culture was grown under the same conditions to an A600 of ~0.6. At this point, methionine biosynthesis was down regulated by the addition of L-lysine, L-threonine, and L-phenylalanine at a final concentration for each of 100 mg/mL and L-leucine, L-isoleucine, and L-valine at 50 mg/mL each. D,L-selenomethionine was added simultaneously to a final concentration of 100 mg/mL. After 15 - 20 minutes, expression of thymidylate kinase was induced by addition of IPTG (isopropyl thio-β-D-galactosidase, Gibco BRL) to 1 mM. Growth of the culture was continued for an additional 3.5 hours until an A600 of 1.5 - 1.6. Cells were then harvested by centrifugation and frozen at -80°C. Under these conditions, the average yield of cell paste was 3.0 to 3.5 g/L.

For protein purification, all buffers were chilled to 4°C prior to use and all procedures were performed at 4°C. Cells (24.8 g wet weight) were resuspended in 125 mL of lysis buffer (25 mM Tris (pH 7.8), 500 mM NaCl, 10% glycerol, 25 mM imidazole, 5 mM 2-mercaptoethanol, 0.2 mg/mL DNAse I) and ruptured by using an American Instrument French Press at 16,000 PSI. The lysate

was clarified by centrifugation at 39,200 X g for 60 minutes in a JA20 rotor. The supernatant was filtered by using a Nalgene 0.2 μm filter unit. The filtered supernatant was applied at 74 cm/hr to a Qiagen NTA Superflow column (1.6 cm i.d. X 11 cm (CV = 22 mL)) charged with nickel that was pre-equilibrated with EQ buffer (25 mM Tris (pH 7.8), 500 mM NaCl, 10% glycerol, 25 mM imidazole, 5 mM 2-mercaptoethanol). The column was washed with 7.7 CV of EQ buffer, 12.5 CV of wash buffer (25 mM Tris (pH 7.8), 500 mM NaCl, 10% glycerol, 50 mM imidazole, 5 mM 2-mercaptoethanol) and eluted with 1.4 CV of elution buffer (25 mM Tris (pH 7.8), 500 mM NaCl, 10% glycerol, 300 mM imidazole, 5 mM 2-mercaptoethanol). During the elution the linear velocity was decreased to 42 cm/hr. The eluted fraction was treated with DTT to achieve a final concentration of 10 mM and dialyzed extensively against nitrogen sparged dialysis buffer (25 mM Tris (pH 7.8), 500 mM NaCl, 10% glycerol, 10 mM DTT, pH 7.8).

The Mono Q analytical run was performed using 50 mL native TMK (14 mg/mL) diluted to 200 mL with 20 mM Tris (pH 8.0). The sample was loaded onto a Mono Q (Amersham Pharmacia Biotech) column equilibrated with 20 mM Tris (pH 8.0) and run through a 20-40% (20 mM Tris (pH 8.0) + 1.0M NaCl) gradient in 40 mL with a flow rate of 1.0 mL/min. The Mono P column run was performed using 50 mL TMK (14mg/mL) diluted to 200 mL with 25 mM bis-Tris (pH 6.71). The sample was injected onto a Mono P column (Amersham Pharmacia Biotech) equilibrated with 25 mM bis-Tris (pH 6.71) and run through a step gradient of 0-100-0 % Polybuffer Mix 96/74 (20:1), pH 5.80. Gel filtration studies were carried out on a Superose 200 column with a 500 mL sample of thymidylate kinase at a concentration of 4.2 mg/mL using 50 mM Tris (pH 8.5), 500 mM NaCl, 5 mM 2-mercaptoethanol, and 0.5% glycerol at a flow rate of 1 mL/min. For dynamic light scattering experiments, samples were mixed in 1.5ml eppendorf tubes, then sterile filtered through a 0.22 mm ceramic membrane (Whatman). 20 mL of solution is read in a quartz cuvette in a Dyna Pro Molecular Sizing Instrument (Protein Solutions, Inc., Charlottesville, VA).

The native protein was exchanged into 50 mM Tris (pH 7.8), 5 mM 2-mercaptoethanol to a concentration of 15 mg/mL and screened for crystallization conditions using Crystal Screen I, Crystal Screen II, and MembFac Screen (Hampton Research, Laguna Niguel, CA). The most encouraging lead was from Hampton Crystal Screen I condition 23: 30% PEG 400, 0.1M Na HEPES pH 7.5, 0.2M $MgCl_2$. Follow up screens indicated that PIPES buffer was most conducive to crystal formation.

The initial crystals of the thymidylate kinase were stacks of small plates that were inseparable and unusable for diffraction studies. Biochemical analysis of the protein revealed that the sample was substantially pure by sodium dodecylsulfate polyacrylamide-gel electrophoresis (SDS-PAGE) analysis, but isoelectric focusing (IEF) gels revealed at least two distinct isoelectric species. It is likely, although yet unproven, that these isoelectric species were the cause of the morphology of the thymidylate kinase crystals. Further efforts at purification with a Mono Q column indicated that separation of these species would be difficult and it was not clear that large scale isoelectrofocusing using a Mono P column or preparative isoelectric focusing would improve the separation because of the small differences in pI. A series experiments exploring the feasibility of preparative isoelectric focusing experiments using PrIME (preparative isoelectric membrane electrophoresis) was hampered due to precipitation of the protein near its pI. Gel filtration did reveal that thymidylate kinase behaves as a dimer in solution, confirming earlier literature reports for the related *E. coli* and yeast TMK. The initial crystallization conditions contained 200 mM $MgCl_2$ and later experiments showed that at least 150 mM $MgCl_2$ was required for crystal formation. Dynamic light scattering experiments in the presence of $MgCl_2$ revealed an interesting phenomenon where protein aggregation was reduced in the presence of $MgCl_2$ over a number of hours leading to a monodisperse, dimeric sample suitable for crystallization.

The stacked plates were eventually transformed into single crystals through iterative streak seeding and crystallization on hanging or sitting drops with thymidylate kinase in 0.1 M PIPES (pH 6.6), 14-19% PEG 400, 0.2 M MgCl2. This

technique involved taking the multinucleated crystals, crushing them into microcrystals, and using a dilution series of this suspension of microcrystals for seeding. It was observed that this second round of crystals were usually less multinucleated than when crystal formation was allowed to proceed via spontaneous nucleation. A second round of streak seeding was usually necessary in order to obtain multiple single crystals. Refinement of the streak seeding technique resulted in native and selenomethionine TMK crystals on the order of about 100 μm x about 100 μm x about 20 μm.

Subsequent crystallization experiments also indicated that a protein concentration of 7 mg/mL was able to yield suitable crystals. The crystallization solution was a cryoprotective agent making it straightforward to freeze the crystals in liquid nitrogen for data collection. Selenomethionine thymidylate kinase was exchanged into 10 mM Tris (pH 7.8), 10 mM DTT and concentrated to 7 or 14 mg/mL for crystallization experiments.

B. X-ray Diffraction Characterization

Thymidylate kinase crystals were generally too small for useful data collection using standard x-ray diffraction equipment. Therefore, all data collection was carried out at the Advanced Photon Source (Argonne, IL). The structure of *S. aureus* thymidylate kinase was determined by multiple anomalous dispersion (MAD) using synchrotron radiation. Crystals were of the space group $P2_1$ with cell constants a = 49.8 Å, b = 90.1 Å, c= 46.5 Å, $\alpha=\gamma=90°$ and $\beta = 101.8°$. The Matthews coefficient for these crystals assuming that there are two molecules in the asymmetric unit is 2.1 Å/Da with 40% solvent. Two MAD data sets were collected B one at 2.7 Å and one at 2.3 Å.

Two selenomethionine multiple anomalous dispersion (MAD) experiments were performed (2.7Å resolution and 2.3 Å resolution) using three different wavelengths (remote wavelength 1.0332 Å, 12000 eV, inflection point wavelength 0.979746 Å, 12654.8 eV, and the peak wavelength 0.979617 Å, 12656.5 eV).

C. Heavy Atom Derivative

Selenomethionine thymidylate kinase was expressed using downregulation of methionine biosynthesis (T.E. Benson et al., Nat. Struct. Biol., 2:644-53 (1995); G.D. Van Duyne et al., J. Mol. Biol. 229:105-24 (1993)) and purified in order to obtain de novo phases by multiple anomalous dispersion (W.A. Hendrickson, Science 254:51-8 (1991)). Anomalous difference Patterson maps revealed six selenium sites (three for each of the two monomers in the asymmetric unit) (Figure 9). Patterson maps at 2.7 Å showed that the atomic positions for the seleniums were not well resolved, but maps at 2.3 Å clearly defined the atomic positions of the heavy atoms. Unfortunately, the MAD phases for data collected at 2.3 Å were of lower quality than the phases at 2.7 Å, so initial model building was performed using the MAD phased map to 2.7 Å (Figure 10). Subsequent refinement was conducted against the 2.3 Å data, and this higher resolution structure is the one reported here.

D. Phase Combination

Each of these individual data sets was indexed and integrated separately (see Tables 7 and 8 for integration statistics). The data sets were scaled to each other using the program SCALEIT in the CCP4 Program Suite (Collaborative Computational Project N4, Acta Cryst. D50:760-3 (1994)). Patterson maps revealed six selenium sites (three for each monomer in the asymmetric unit) whose locations were determined by direct methods using SHELX (G.M. Sheldrick & R.O. Gould, Acta Cryst. B51:423-31 (1995)). Heavy atom refinement and phase calculations were conducted using SHARP (E. La Fortelle et al., A Maximum-Likelihood Heavy-Atom Parameter Refinement and Phasing Program for the MIR and MAD Methods, P. Bourne & K. Watenpaugh, eds., Crystallographic Computing 7 (1997)). Phases calculated in SHARP were solvent flattened using the program SOLOMON (Collaborative Computational Project N4, Acta Cryst. D50:760-3 (1994)) and gave a significantly improved electron density map.

Table 7. Data collection and phasing statistics for structure of *S. aureus* TMK

	λ 1.0332 Å (12000 eV)	λ 0.979746 Å (12654.8 eV)	λ 0.979617 Å (12656.5 eV)
Resolution	2.7 Å	2.7 Å	2.7 Å
No. observations	76,132	62,273	76,145
No. unique refl.	10,901	10,941	10,928
% completeness	100%	100%	100%
R_{sym}	0.085	0.103	0.106
R_{cullis} acentrics	--	0.61	0.67
R_{cullis} anomalous	0.98	0.78	0.69
Phasing power			
centrics	--	1.28	1.21
acentrics	--	2.30	1.83
Mean figure of merit (to 2.7 Å resolution)			
before solvent flattening		0.51	
after solvent flattening		0.94	

Table 8. Data collection and phasing statistics for structure of *S. aureus* TMK

	λ 1.0332 Å (12000 eV)	λ 0.979746 Å (12654.8 eV)	λ 0.979617 Å (12656.5 eV)
Resolution	2.3 Å	2.3 Å	2.3 Å
No. Observations	76,712	123,553	123,372
No. unique refl.	17,661	17,887	17,991
% completeness	98.2%	99.4%	99.3%
R_{sym}	0.083	0.107	0.099
R_{cullis} acentrics	--	0.56	0.61
R_{cullis} anomalous	0.99	0.69	0.70
Phasing power			
centrics	--	1.34	1.38
acentrics	--	2.22	2.04
Mean figure of merit (to 2.3 Å resolution)			
before solvent flattening		0.57	
after solvent flattening		0.87	

E. Model Building and Refinement

At this stage in the structure solution, the coordinates for *E. coli* thymidylate kinase greatly aided the process of model building for placement of the main chain backbone. Model building was done using the program CHAIN (J.S. Sack, Journal of Molecular Graphics 6:224-5 (1988)) and LORE (B.C. Finzel, Meth. Enzymol. 277:230-42 (1997)). Refinement was carried out with XPLOR98 (A.T. Brunger, X-PLOR version 3.1: A system for X-ray Crystallography and NMR, New Haven: Yale Univ. Press, (1992)) incorporating bulk solvent correction during the refinement (J.S. Jiang & A.T. Brunger, J. Mol. Biol. 243:100-15 (1994)). Progress of the refinement was monitored by a decrease in both the R-factor and Free R-factor.

Table 9. Refinement Statistics for structure of *S. aureus* TMK

	R-factor	Free R-factor	No. of reflections
20-2.3 Å F≥ 2σ	0.2366	0.3084	15,908
		Bonds (Å)	Angles(°)
r.m.s deviation from ideal geometry		0.008	1.32
	Number of atoms	Average B-factor	
Protein	2978	27.2	
Waters	174	38.9	
Total	3152	27.81	

Stereochemistry of the model was checked using PROCHECK (R.A. Laskowski et al., J. Appl. Cryst. 26:283-91 (1993)) revealing no residues in disallowed regions of the Ramachandran plot. Figure 9 was made using SETOR (S.V. Evans, J. Mol. Graphics 11:134-8 (1993)) and Figures 3a, 4 were produced in MOLSCRIPT (P. Kraulis, J. Appl. Cryst. 24:946-50 (1991)) and Raster 3D (E.A. Merritt & M.E.P. Murphy, Acta Cryst. D50:869-73 (1994)) while Figures 5a and 6a were produced in MOLSCRIPT (P. Kraulis, J. Appl. Cryst. 24:946-50 (1991)) alone.

F. Assays.

Binding assays to determine if a compound actually interferes with *S. aureus* thymidylate kinase can also be performed. For example, thymidylate kinase activity can be measured by coupling the formation of ADP and TDP to the reactions catalyzed by PD, LDH, and NDP-Kinase, as shown below. Oxidation of NADH is accompanied by a decrease in absorbance at 340 nm, which is measured spectrophotometrically.

1) ATP + TMP $\xrightarrow{\text{Thymidylate Kinase}}$ ADP + TDP

2) ATP + TDP $\xrightarrow{\text{NDP-Kinase}}$ ADP + TTP

3) 2 ADP + 2 PEP $\xrightarrow{\text{PK}}$ 2 ATP + 2 PYRUVATE

4) 2 PYRUVATE + 2 NADH $\xrightarrow{\text{LDH}}$ 2 LACTATE + 2 NAD^+

The standard reaction conditions employed during the kinetic characterization of the enzyme were: 50 mM HEPES, pH 8.0, 50 mM KCl, 2 mM $MgCl_2$, 4 U/ml PK, 5 U/ml LDH, 2 mM PEP, 1.5 mM ATP, 5 U/ml NDP-Kinase, 1.0 mM TMP, 0.22 mM NADH, and 0.8 μg/ml T. kinase. All of the reagents except the T. Kinase were added to a cuvette and mixed, and the mixture was incubated at 24.5°C for 2 minutes. To start the reaction, the T. Kinase was added, the contents of the cuvette were mixed, and the decrease in absorbance at 340 nm was monitored for 4-5 minutes.

SEQUENCE LISTING FREE TEXT

SEQ ID NO:1	recombinant *S. aureus* thymidylate kinase (with polyhistidine [His_6] sequence tag)
SEQ ID NO:2	*E. coli* thymidylate kinase
SEQ ID NO:3	*S. cerevisiae* thymidylate kinase